

PAP JÓZSEF

Relációs adatbázisok felhasználási lehetőségei a történeti kutatásokban*

Az elmúlt években végzett kutatásaim során megismert módszertani eljárás, az adatbáziskezelés rövid ismertetésére vállalkozom ebben az írásban. Konkrét példák segítségével fogom illusztrálni azt, hogy milyen eredményeket adott számomra és kutatótársaim számára a relációs adatbázis használata, és az eredmények mellett milyen problémákkal találtam magam szembe. A bevezető részben először arról fogok írni, hogyan is jutottam el az adatbázisok alkalmazásáig, majd ezt követően az adatbázisépítés kezdeti lépéseiről szólok. Írásommal egy olyan innovatív módszerre szeretném felhívni a figyelmet, melyet nem csupán a kutatói munkámban tudok hasznosítani, hanem tan-zsék- és intézetvezetőként, szakfelelősként is segíti a hétköznapi tevékenységemet.

Vármegyei tisztikar – passzív ellenállás, társadalomtörténeti vizsgálat strukturált adathalmaz alapján

Pályám kezdeti szakaszán a neoabszolutizmus korának vármegyei hivatalnokaival foglalkoztam, táblázataimat az akkor még modernnek számító módszerrel, a Microsoft Excel programban alakítottam ki. A vármegyei tisztviselők egyes életpályaelemeit külön-külön vizsgálva kíséreltem meg leírni az adott közösséget. A kutatás legfontosabb eredményeit 2003-ban monográfiában foglaltam össze.¹ Az ekkor elemzett 6 vármegye hivatalnoki állományát közel 10 000 fő alkotta. A forrásfeltárás során egyenként rögzítettem a vármegyék hivatali kimutatásait, majd ezekből a listákból hoztam létre egy olyan táblázatot, melyben személyenként voltak megtalálhatók az egyes hivatali állomáshelyek és beosztások. A listák összeállításához az alkotmányos időszakok esetében az adott megyék vá-

* A tanulmány elkészítését az EFOP-3.6.1-16-2016-00001 „Kutatási kapacitások és szolgáltatások komplex fejlesztése az Eszterházy Károly Egyetemen” című pályázat támogatta. Ugyancsak köszönöm az NKFIH 112429 „A dualizmus kori magyar országgyűlések tagjainak feltárása és társadalomtörténeti elemzése” címet viselő pályázatának támogatását.

1 Pap 2003.

lasztási jegyzőkönyveit használtam fel, míg az 1850-es évek és a provizórium feltárása során a nyomtatásban megjelent hivatali listákra támaszkodtam. Az innen kinyert adatokat egészítettem ki az országos kormánysszervek iratanyagából származó információkkal. A legfontosabb kútfő, a Magyar Nemzeti Levéltár D 188-as fondjában található személyi kimutatási kartonok irategyüttese volt.² Az egyének beazonosítása után rögzítettem azok származását, születési és halálozási helyét, képzettségi szintjét és a karrier egyes állomásait. Az adatok forrását soronként jelöltem meg, ezáltal minden egyes információ visszakereshetően került feldolgozásra. Mivel a származásra vonatkozó adataim hiányosak voltak (az 50-es évek nyilvántartásait leszámítva), az egyének többségének nemesi címére csak következtetni lehetett. Az adatgyűjtés időszakában még nem álltak rendelkezésre azok az elektronikus adatbázisok, melyek segítségével ma már sokkal egyszerűbb egy-egy személy származásának a felderítése.³ Hasonlóan hiányosak voltak a halálozási adatok is, hiszen ezeket a hivatali kartonok csupán a legkritikább esetben tartalmazták. A feldolgozott életrajzok mennyisége azonban nem tette lehetővé egy olyan általános vizsgálat elvégzését, mely során a manuálisan kutatható genealógiai szakirodalom, a helyi sajtó és a gyászjelentések szisztematikus feldolgozása megtörténik. Ma, a kutatás kezdő időszakát számolva, huszonöt év múltán, sokkal egyszerűbb az ilyen típusú adatok összegyűjtése.⁴

Az összeállított anyag, mely megynként, alfabetikus sorrendben tartalmazza a tisztviselőket, maga félezer oldalt tett ki. Nyomtatásban természetesen nem jelenhetett meg, kéziratban maradt. A névtár feldolgozását egyszerű adattáblák segítségével végeztem el. Az életrajzi elemeket kóddal láttam el, külön kódot kaptak az egyes, politikailag jól definiálható korszakokban történt hivatalviselésre utaló adatok. A kódolás során elválasztottam egymástól a reformkori, az 1848–49-es, a szabadságharc alatti osztrák megszállást követő, az 1849 és 1854 közötti, az 1854 és 1860 közötti, az 1860–61-es, az 1861–1867 közötti, valamint az 1867 utáni életpályaszakaszokat. Ebből egy táblázatot képeztem, és ebben kerestem az azonos mintázattal rendelkező személyeket.

2 Az általam felhasznált kimutatások időközben a Magyar Nemzeti Levéltár digitalizációs programjába is bekerültek, Tuza Csilla vezetésével készül egy digitális adatbázis, mely az MNL által fejlesztett Névtérben szándékozik megjeleníteni a minősítési kartonok adatait. Tuza 2016: 1347–1349.

3 A genealógiai adatgyűjtés kiváló kiindulópontjai voltak az Arcanum Kft. által megjelentetett munkák. 2013-ban látott napvilágot a Család- és helytörténeti irodalom c. blu-ray formátumú gyűjtemény, melyet merevlemezre másolva is meg lehetett a kiadótól vásárolni. A lemezen 140 kötetnyi családtörténeti irodalmat adtak közzé. Ezt egészítette ki a szintén 2013-ban megjelent Családtörténeti folyóiratok c. DVD. Sajnos azonban ma már ezek a lemezek, adatállományok nem vásárolhatók meg, az online tartalomszolgáltató oldalakon pedig nincs fent a 2013-ban kiadott teljes anyag.

4 Legegyyszerűbben az Országos Széchényi Könyvtár Gyászjelentés-gyűjteményét lehet felhasználni erre a célra. Korábban ez a gyűjtemény csupán a <https://www.familysearch.org> oldalon volt elérhető, ahol alfabetikus sorrendben, gyakorlatilag átlapozással lehetett benne keresni. Az Országos Széchényi Könyvtár készített azonban egy saját keresőrendszert, mely sokkal felhasználóbarátabb felületet teremtett. (<https://dspace.oszk.hu/handle/20.500.12346/663648>) Nagy előnye ennek az oldálnak, hogy nemcsak az elhunyt nevére lehet benne keresni, hanem a családtagok között is megtalálhatjuk a keresett nevet. Ezáltal a családi kapcsolatok feltárását nagymértékben megkönnyíti. Fontos kiemelni még – az interneten fellelhető számtalan genealógiai gyűjtemény közül – a Magyar Családtörténet-kutató Egyesület honlapját, ahol szintén több adatbázis áll a kutatók rendelkezésére.

	Reformkor	1848–49	1849	1849–54	1854–60	1860–61	1861–65	1867–
személy1								
személy2								
személy3								

1. ábra A vármegyei hivatali elit életpálya-elemzésének táblázata

Az egyes vármegyék különböző időszakban szolgálatban álló hivatalnoki karát ennek alapján írtam le. Elsősorban a kontinuitás-diszkontinuitás mértékét próbáltam meghatározni. Vizsgálatom ezen túl arra koncentrált, hogy ez a hivatalnoki kör milyen képzettséggel, előélettel és további szolgálattal rendelkezett. Az adott eszközökkel erre nyílt lehetőség, erre – a mai szemmel már egyszerű – kérdésre tudtam választ adni, és kimutatni viszonylag jelentős kontinuitást ott, ahol a passzív ellenállás elméletéből kiindulva inkább annak ellenkezőjét kellett volna tapasztalni.⁵ Ez is egy jelentős kutatási eredmény volt, azonban azt hiszem, a feldolgozott adatokból a jelenleg rendelkezésre álló informatikai eszközök és módszerek segítségével ma már sokkal összetettebb képet tudnék felrajzolni. A továbblépéshez azonban módszertani váltásra van szükség.

5 Az elvégzett elemzések tükrében megállapíthattam, hogy a „Bach-huszár” fogalom mögött elsősorban magyar származású személyeket kell értenünk, ők azonban más vármegyéből áthelyezetteként szolgálati helyeiken idegennek minősültek. Az elemzett vármegyékben megjelennek ezek a „belföldi” idegenek. Az elemzett 6 vármegye nemcsak befogadó, hanem kibocsátó is volt egyben. A kinevezési gyakorlatban a hajdani alkotmányos szerepvállalás csak másodrangú szempontnak számított, hiszen ekkor 1848-as hivatalnokok is állásokhoz jutottak, a vezető pozíciókban pedig meghatározó maradt az 1849 előtti szolgálati múlttal rendelkezők szerepe. 1854 után azonban a tisztviselők egy jelentős része elvesztette állását, ennek okai közül a pályáztatás során támasztott magas szakmai követelményeket (felsőfokú jogi vagy közigazgatási végzettség, német nyelv ismerete) és a hagyományos léttértől távoli munkahely miatti elköltözés szükségességét kell kiemelni. Az 1860–61-es kurzusváltás törést jelentett ugyan a neoabszolutizmus kori hivatalviselők számára, de ez korántsem volt általános érvényű tendencia. A „Bach-huszárok” számára az időleges egzisztenciális válságot az önkényuralom újabb korszaka oldotta meg. A Schmerling-provizórium apparátusát elsősorban új hivatalnokok töltötték be, de a vezető pozíciókban helyet kaptak a helyi származású, vagy az adott vármegyében letelepedett, ötvenes években szolgált magyarok. Mellettük kisebb-nagyobb szerepet kaptak az 1860–61-es hivatalnokok is. Idegen ajkú ember azonban elvéve akad köztük (a feldolgozott 6 vármegyében egyetlen ilyen tisztviselő található). 1867 újabb jelentős határkö volt a szakapparátus számára, általános érvényű megállapítást itt sem lehet tenni, hiszen nagy eltérés figyelhető meg az egyes vármegyék kinevezési gyakorlataiban. A kutatás eredményével kapcsolatban lásd: Pap 2003: 274–279.

Két (vagy több) választás Magyarországon – társadalomtörténeti vizsgálatok új eszközökkel

A vármegyei tisztikart feldolgozó munkákat követően 2004-ben a dualizmus kori népképviselőkkel kezdtem el foglalkozni. Ebben a ma is folyó kutatásban több új módszertani elemet használtam fel. Az adatrögzítés ma már relációs adatbázisban történik, a kvantifikálható jelenségeket a statisztika eszközszerével is vizsgálom, digitális térképek segítségével ábrázolok bizonyos jelenségeket, valamint elkezdtük a hálózatelméleti megközelítés lehetőségeinek a feltárását is.⁶ A kutatási folyamatnak fontos részét képezi továbbá a keletkezett vizsgálati eredmények vizuális megjelenítése és ennek módszertani átgondolása. Olyan, nem öncélú ábrázolást jelent számomra, mely valamilyen pluszt lehetőséget rejt magában, nemcsak segíti a szövegben található információk megértését, hanem további tudományos felismerések kiindulási alapjául szolgálhat. Saját kutatásaim során jelenleg elsősorban a térinformatika és a digitális térképek területéről tudok ilyen példákkal szolgálni.⁷ Azonban jelen kötetben erről a kérdésről külön tanulmány szól, írásom következő részében az adatrögzítés területére fogok koncentrálni.

A. Adatrögzítés relációs adatbázisban

2004-ben, amikor a dualizmus kori népképviselőkkel kezdtem el foglalkozni, szintén az Excelt, valamint az akkor megismert SPSS programot használtam az adattárolásra. A témában megjelent első komolyabb cikkemben az SPSS programra mint adatbáziskezelő szoftverre hivatkoztam. A vizsgálat alapját képező táblázatról pedig a következőket írtam: „Adatbázisunkban 600 képviselőre 34 különböző típusú információ található.”⁸ Ez a fogalomhasználat alapvetően megfelelt a hazai szakirodalom hagyományának, ahol az adatbázis alatt az esetek nagy többségében, egy-egy adattáblát értenek a kutatók.

2007-ben azonban új alapokra helyeződött a munka, melynek központi eleme éppen az adatbázis fogalom tartalommal való megtöltése volt. Az önmagukban álló, egy-egy publikáció elkészítésére használható adattáblák helyét egy MS Access programon belül elkészített relációs adatbázis vette át. A módszerrel az akkor még hallgató Ballabás Dániel ismertetett meg. Innentől kezdve gyakorlatilag együtt dolgozunk, az adatbázis alapelveit is egymással egyeztetve alakítottuk ki. Ezt a váltás jelezte a 2007-es cikkem tanulmánykötetben megjelent átdolgozott és kiegészített szövege is. 2014-ben az SPSS-re már csupán adatelemző szoftverként hivatkoztam. A 6. lábjegyzetben

⁶ Ballabás 2019: 33–67.

⁷ Pap 2019: 163–189.; Rózsa 2019: 103–133

⁸ Pap 2007: 6.

pedig az alábbi szöveg olvasható: „A tanulmány elkészítésekor az adatbázis fogalmán egy adattárat (Excel-táblázatot) értettünk. Épp az ekkor tapasztalt – adatrendszerelési, rögzítési és feldolgozási – problémák késztettek minket arra, hogy új munkamódszert keressünk, és a valódi adatbázis-kezelés eszközét (MS Access adatbázis-kezelő program) alkalmazzuk kutatásaink során.”⁹ Miért is volt szükség erre a kiegészítésre? Miben jelent többet egy relációs adatbázis egy egyszerű adattáblánál? Ezt próbálok a következőkben röviden bemutatni. A fejezet hátralévő részében az adatbázisépítés elemi lépéseiről fogok szólni.¹⁰

1. Adatbázisépítési alapok történészszemmel

Az adatbázis alapeleme tehát az egyszerű adattábla, melyben sorokat és oszlopokat különböztetünk meg. A sor maga a vizsgálat alapeleme, melyet rekordnak nevezünk. Az oszlopok – a mezők –, melyeket a statisztika a változók fogalommal illet. A sorok és oszlopok kereszteződésében található cellákat pedig mezőértéknek nevezzük. Egy adattáblában a rekord például maga az országgyűlési képviselő, rá vonatkozó mező (változó) lehet a nem, a mezőérték pedig az adott személy neme.

	Mező 1.	Mező 2.
Rekord 1.	mezőérték	
Rekord 2.		
Rekord 3.		

2. ábra Adattábla egyszerűsített formája

Ez a struktúra alapvetően megfelel az Excelből ismert táblának, azonban egy adatbázisban számtalan táblát helyezhetünk el, melyben különböző információkat tárolhatunk. Ezek a táblák önmagukban is értelmezhetők lehetnek, de egymás változóiként is felhasználhatjuk őket. Maradva az előző példánknál, változó lehet például a születés helye, mely rekordonként, tehát személyről személyre haladva más és más mezőértéket kaphat. De a település lehet egyben rekordnév is egy másik táblában, ahol a rá vonatkozó egyéb változók kerülnek rögzítésre. Ilyenek lehetnek például a közigazgatási adatok (megye, vagy járás, választókerület), a demográfiai jellemzők, a földrajzi koordináták, stb. Az adott rendszer tehát a hely szempontjából két alapvizsgálatra is lehetőséget ad, elemezhetjük a képviselőket születési hely alapján, de ha ismerjük a települések demográfiai adatait, ugyanabból az adatbázisból akár a településekre vonatkozó kérdésekre is választ kaphatunk. Az adott táblánkat csupán egyszer kell elkészítenünk, azonban egymással szabadon variálva többször is felhasználhatjuk azokat.

⁹ Pap 2014: 21.

¹⁰ Az adatbázis konkrét kutatási folyamatban való felhasználásáról lásd Nagy Andor tanulmányát e kötetben: Nagy 2019: 69–101.

Az alábbiakban bemutatom azt, hogy miként áll össze egy születési és halálozási adatokat tartalmazó egyszerű táblastruktúra, melyben az adott településekről elemi közigazgatás-történeti információkat is tárolunk.

Első lépésben elkészítjük a „Személy” táblát, melyben minden egyes ember külön azonosítót kap. A táblát úgy kell elképzelni, mint egy egyszerű Excel-táblát, ahol a soroknak és oszlopoknak nevet adunk. Gyakorlatilag ennek megfelelően működik az adatbáziskezelő programok táblaszerkesztő modulja is, csak ott az adat típusát¹¹ is be kell állítanunk. A „Személy” táblában külön azonosítóval, ID-vel, egy nullától különböző számmal jelölölünk minden egyes embert, és a következőkben az adott személyre mindig ezzel a számmal hivatkozunk. Minden személynek tehát, hasonlóan a személyi számhoz vagy az adóazonosító jelhez, egyedi azonosítót adunk. Külön mezőben tároljuk továbbá a nevet, a születési és a halálozási helyet. A településeknek azonban már nem a nevét írjuk be, hanem az azonosítóját (lásd alább). Az oszlopok (mezők vagy változók) száma pedig szabadon növelhető. Erre a táblára azonban csupán olyan adatokat célszerű felvennünk, melyből egy-egy jellemzi az adott személyt. Szerepelhet például az apa és az anya (akik egyébként maguk is külön ID-vel ellátott személyek lehetnek ugyanebben a táblában). A házastársat vagy a gyerekeket már nem itt rögzítjük, hiszen belőlük több is lehet, így egy cellába változó számú személyt kellene feltüntetnünk.

Személy ID	Név	Születési hely	Halálozási hely	Apa	Anya
1	Személy 1.	1	1	2	3
2	Személy 2.	1	2	9999	9999
3	Személy 3.	2	2	9999	9999
8888	Nem vonatkozik rá	8888	8888	8888	8888
9999	Ismeretlen	9999	9999	9999	9999

3. ábra A Személy tábla szerkezete

A fenti példán látható, hogy a „Személy” táblában az 1-es ID vel jelölt személynek a kettes az apja, a hármas pedig az anyja. Külön felvettünk egy ismeretlen személyt jelző adatot (9999), mely kód megjelenik az apánál és az anyánál, mivel a szüleit esetükben nem ismerjük. Az ismeretlen személynek természetesen ismeretlen a születési helye is. Szükségünk lesz még egy másik speciális „személyre”, melyikkel azt jelöljük, hogy az adott elemhez nem tartozik senki, tehát nem ismeretlen emberről van szó, hanem arról, hogy ilyen személy nem létezett. Fontos arra is figyelni, hogy minden cellát

¹¹ A rögzített adatunk lehet szám, dátum, rövid vagy hosszú szöveg. Rögzíthetünk pénznemeket és URL hivatkozásokat, vagy akár egy automatizált számítást is tartalmazhat az adott mező.

ki kell töltenünk, hiszen az is adatot jelent számunkra, ha valamiről biztosan tudjuk állítani, hogy létezett, azonban pontos személye nem ismert. Egy házastárs esetében például mást jelent az ismeretlen személy, és mást a „nem vonatkozik rá adat”, utóbbi használata ugyanis arra utal, hogy biztosan tudjuk, az adott embernek nem volt házastársa. Adhatunk azonban például külön kódot a még nem kutatott adatoknak is, mely a későbbiekben jelölheti számunkra azt, hogy kivel kapcsolatban kell még információkat gyűjtenünk.

Most nézzünk meg egy másik példát, a települések rögzítési lehetőségeit. Településekre vonatkozó adatokat rendszeresen tárolunk hagyományos táblázatainkban, általában azonban a neveket tüntetjük fel az adott tábla megfelelő cellájában. Egy adatbázisban azonban másként célszerű dolgozni, itt külön táblában tároljuk el a települések adatait, mely egyben a „Személy” tábla önálló adatforrása is, hiszen ott már a települések azonosítói jelennek csupán meg a megfelelő cellákban. Ahhoz tehát, hogy a „Személyek” táblában ki tudjuk tölteni a születési és halálozási helyekre vonatkozó adatokat, a „Település” táblában már adatoknak kell szerepelni. A „Település” táblához hasonló táblákat tekinthetjük akár szótáraknak¹² is, melyekben egyedi azonosítóval ellátva jelenítjük meg az adott változóértékeket, jelen esetben a településeket. A szótáraink egy részét az adatbázis-építés elején fel tudjuk tölteni adatokkal. Ilyen lehet például a biológiai nemet tartalmazó egyszerű szótár (1: férfi; 2: nő), vagy a vármegyék szótára. Más esetekben nagyobb előmunkálatok szükségesek ehhez. A települések tipikusan ilyenek, hiszen a vizsgált terület, ami lehet egy járás, egy vármegye vagy akár egy ország esetében nagyszámú helység fordul elő. A dolog csak bonyolódik, ha több időmetszetet veszünk figyelembe, hiszen ebben az esetben nemcsak a névvariációkkal kell megküzdenünk, hanem az összeolvadás vagy szétválás is bonyolítja a helyzetünket. Nem minden esetben tudunk azonban minden elemet előzetesen rögzíteni, lehetőségünk van arra is, hogy szótárunkat folyamatosan fejlesztjük, és akkor vegyünk fel bele egy új rekordot (ebben az esetben a településnevet), ha az előfordul a feldolgozott források között. Ebben a táblában is el kell helyoznünk az Ismeretlen és a Nem vonatkozik rá adatot.

Település ID	Település név	Járás/RTV/TJV 1848	Járás/RTV/TJV 1876
1	Eger	1	1
2	Szolnok	2	2
8888	Nem vonatkozik rá	8888	8888
9999	Ismeretlen	9999	9999

4. ábra A Település tábla szerkezete

12 A szótárak alatt egy olyan adattáblát értünk, mely egy adott változó értékeit tartalmazza. Itt tüntetjük fel a kódokat és azok feloldását. Más táblákban pedig ezeket a kódokat használjuk. A szótár fogalma tehát ebben az értelemben a kód megfejtését tartalmazó táblát takarja.

Az fenti „Település” táblán jól látható, hogy az egyrészt egyszer szótárként funkcionál, hiszen a „Személy” táblában szereplő születési és halálozási helykódok feloldása itt található, másrészt azonban ugyanez a tábla adatokat is rögzít magáról a településről. Példaként két időmetszet járási azonosítóját helyeztem el egy-egy mezőben, de itt feltüntethetnénk akár egy adott népszámlálás vonatkozó községsoros adatait is.

A települések rendszerszerű rögzítéséhez azonban szükségünk van egy „Járás” táblára is, ahol a közigazgatási beosztást rögzítjük. A „Járás” tábla tehát ismét szótár és adat-tábla is egyben. Itt jeleníthetjük meg például egy adott időszak szolgabíróját is, ebben az esetben a „Személy” tábla jelenti a szótárt, hiszen onnan vesszük az egyén azonosítóját. A járási beosztás esetében szemléltetni lehet a „Nem vonatkozik rá” adattípus felhasználhatóságát is. A táblázatban főszolgabírókat rögzítünk, Eger rendezett tanácsú város esetében azonban ilyen funkciót nem töltött be senki, hiszen a város a dualizmus időszakában rendezett tanáccsal, tehát járási joggal rendelkezett ugyan, élén azonban polgármester, nem pedig főszolgabíró állt. Ez jelen esetben egy triviális információ, azonban bonyolultabb kérdéseknél nagyon fontos lehet az ilyen egyedi jellemző ismerete.

Járás ID	Járás/RTV/ TJV név	Vármegye 1848	Vármegye 1876	Főszolgabíró 1876.01.01.
1	Eger rtv	1	2	8888
2	Szolnoki	1	3	1
8888	Nem vonatkozik rá	8888	8888	8888
9999	Ismeretlen	9999	9999	9999

5. ábra A Járás tábla szerkezete

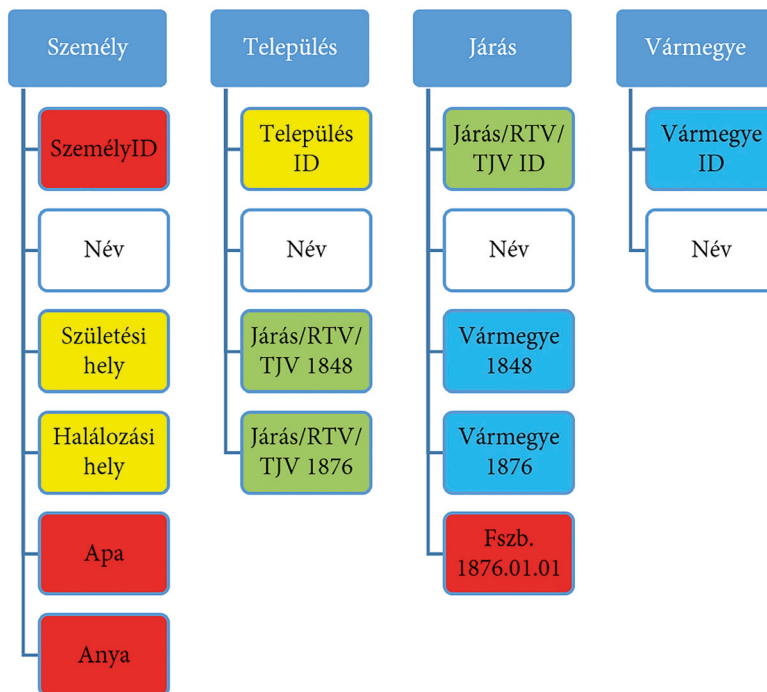
A példánkban szereplő struktúrában a „Vármegye” tábla már egy egyszerű szótár-tábla, melyben minden elem önállóan szerepel, tekintet nélkül azok időbeliségére és az egymás közti esetleges területi átfedésekre.

Település ID	Vármegye név
1	Heves és Külső Szolnok
2	Heves
3	Jász-Nagykun-Szolnok
8888	Nem vonatkozik rá
9999	Ismeretlen

6. ábra A Vármegye tábla szerkezete¹³

13 A táblázat az úgynevezett adattáblanézetben mutatja be a példaként szerepeltetett „Vármegye” táblát. Minden egyes tábla elkészítésekor azonban a tervezőnézetet használjuk, itt állítjuk be az egyes oszlo-

A fent bemutatott táblák struktúrája tehát a következőképpen néz ki:



7. ábra Mintaadatbázis táblastruktúrája

Az ábrán azonos színekkel jelöltem azokat a mezőket, melyek egymással kapcsolatba hozhatók, és ezek segítségével jöhet létre adattábláink között a relációs kapcsolat. Ezek a kapcsolatok is többfélék lehetnek. Egy az egyhez kapcsolatáról beszélünk, ha az egyik tábla egy eleméhez a másik táblának pontosan egy eleme kapcsolódhat. Erre nagyon nehéz példával szolgálni a történelem területéről, hiszen ez valójában egy függvényyszerű kapcsolatot¹⁴ jelent. Beszélhetünk egy a többhöz kapcsolatáról, ilyen például az apa-gyerek kapcsolat, hiszen az egyik esetben egy szereplő jelenhet meg csupán (mindenkinek egy vér szerinti apja van), azonban egy apához több gyermek is kapcsolódhat. Létezik még több a többhöz típus is, ilyen például a házastársi viszony. Itt egy feleséghez több férj és egy férjhez több feleség is kötődhet.¹⁵

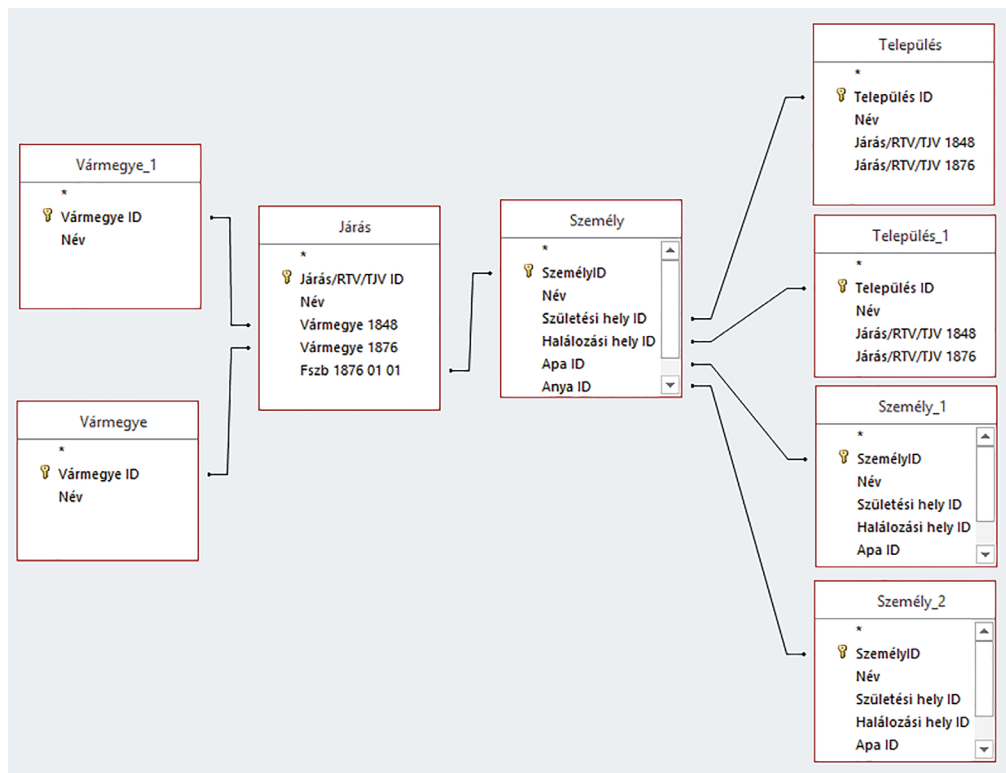
pok tulajdonságait, elnevezzük azokat (Mezőnév), megadjuk adattípusukat (Adattípus: hosszú/ rövid szöveg, szám, számláló (sorszámozott), dátum, idő, igen/nem, hivatkozás, melléklet, számított) és alkalmazhatunk egyértelműsítő leírásokat.) Az adattípus megadásakor lehetőségünk van alapértelmezett értékeket beállítani, használhatunk ún. beviteli maszkokat, melyekkel egyrészt megkönnyíthetjük az adatbevitelt, másrészt elkerülhetjük a téves adatok rögzítését.

14 Függvényyszerű kapcsolat esetén az egyik ismérv szerinti hovatartozás egyértelműen meghatározza a másik ismérv szerinti hovatartozást, tehát egyik jelenségből 100%-os bizonyossággal következik a másik.

15 A genealógiai kapcsolatok adatbázisba rögzítéséről bővebben lásd Ballabás Dániel tanulmányát e kötetben: Ballabás 2019: 33–67.

2. Adatok új köntösben, a lekérdezések

A kapcsolatok irányát és minőségét az adattáblák egymáshoz rendelésekor, az úgynevezett lekérdezések elkészítésekor kell megadnunk.¹⁶ Ugyanis a fenti táblastruktúra alapján különböző kérdésekre kereshetnünk választ. Meg szeretnénk találni például a járások főszolgabíróit, akikről szeretnénk tudni, kik voltak a szüleik, maguk hol születtek és hol haltak meg. Ez a kérdést az alábbi kapcsolatok segítségével tudjuk megválaszolni:



8. ábra Lekérdezés a mintaadatbázisban

¹⁶ Gyakran találkozunk olyan adatbázisokkal, ahol ezeket a kapcsolatokat előzetesen létrehozzák. Ezek azonban általában statikus rendszerek, ahol a lekérdezéseket nem a felhasználók alakítják ki, hanem előzetesen állítják azokat be. Ezen az elven működik például a felsőoktatásban szinte mindenki által ismert Neptun (ETR) rendszer is. A mi állományunk azonban sokkal dinamikusabb annál, hogy ilyen előzetes beállításokkal megelégedjünk.

Jelen tanulmányban nem tudok kitérni a lekérdezések minden típusára, hanem csupán a legegyszerűbb formát, a választó lekérdezést mutatom be. Csak felhívom a figyelmet arra, hogy a lekérdezések segítségével korábbi adatainkból új táblázatokat tudunk létrehozni (táblakészítő lekérdezés), összerakhatunk különböző adatsorokat (hozzáfűző lekérdezés), frissíthetjük korábbi táblázatunkban található adatokat (frissítő lekérdezés), törölhetünk adatokat korábban készült táblákból (törölő lekérdezés), keresztáblákat hozhatunk létre (keresztáblás lekérdezés). Különösen hasznosak a lekérdezésekkel véggezhető számítások. Kivonhatunk például egymásból értékeket, megállítva ezáltal a vizsgált személyek életkorát, összesíthetünk adatokat, megadva ezzel például egy életpálya alatt eltöltött iskolaéveket.

A képen szereplő, az Access programban létrehozott válaszon, lekérdezésen jól látható, hogy egy-egy táblát többször is fel tudunk használni. A grafikus felületen rendkívül egyszerűen, az egérmutató segítségével hozhatunk létre kapcsolatokat a különböző táblázatok sorai között. Ügyelni kell azonban arra, hogy az összekötött elemek adattípusa azonos legyen. A kapcsolat ugyanis csak ebben az esetben működik. Tehát például minden ID típusú adatnál a tábla létrehozásakor a „szám” adattípust állítottuk be. Ha ezt elvettük, később is korrigálhatjuk, ez azonban adatvesztéssel is járhat. Ha a lekérdezést lefuttatjuk, akkor a „Személy” táblából eredményül kapjuk a főszolgabírókat, hiszen csak az ő azonosítójuk szerepel a járás tábla megfelelő mezőjében. Ezeknek a személyeknek a szülei szintén a „Személy” táblából kerülnek kiválogatásra az apa és az anya ID alapján. A „Település” táblát kétszer használtam fel, és így tettem a vármegyével is. A lekérdezés eredménye a következő táblázat lesz.

Járas	Vármegye 1848	Vármegye 1876	Név	Születési hely	Halálozási hely	Apa	Anya
Szolnoki	Heves és Külső Szolnok	Jász- Nagykun- Szolnok	Személy 1	Eger	Eger	Személy 2	Személy 3
Eger rtv.	Heves és Külső Szolnok	Heves	Nem vonatkozik rá	Nem vonatkozik rá	Nem vonatkozik rá	Nem vonatkozik rá	Nem vonatkozik rá

9. ábra A főszolgabírókra vonatkozó lekérdezés eredménytáblája

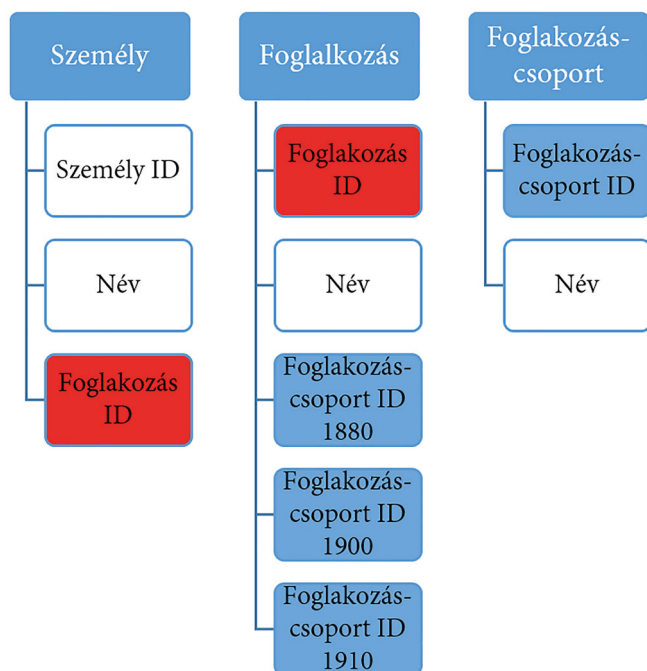
Ha azonban minden személyem születési és halálozási adatára és szüleire lennék kíváncsi, akkor ki kell vennem a lekérdezésből a járásokat és a vármegyéket (vagy legalábbis a példán belül a születési és halálozási helyhez kell kapcsolnom azokat, mi által egyben megszüntetem a járás és a személyek között fennálló kapcsolatot). Eredményem ebben az esetben egy másik táblázat lesz, melyben a paramétereknek megfelelő személyek kívánt adatai jelennek meg. Érdeemes megfigyelni, hogy ezekben a lekérdezésekben azokkal a korábbi táblákkal szemben, melyekből a lekérdezést végrehajtjuk, már nem a kódok jelennek meg, hanem azok feloldása látható.

Név	Születési hely	Halálozási hely	Apa	Anya
Személy 1	Eger	Eger	Személy 2	Személy 3
Személy 2	Eger	Szolnok	Ismeretlen	Ismeretlen
Személy 3	Szolnok	Szolnok	Ismeretlen	Ismeretlen
Ismeretlen	Ismeretlen	Ismeretlen	Ismeretlen	Ismeretlen
Nem vonatkozik rá	Nem vonatkozik rá	Nem vonatkozik rá	Nem vonatkozik rá	Nem vonatkozik rá

10. ábra A személyekre vonatkozó lekérdezés eredménytáblája

3. Bonyolultabb táblastruktúra létrehozása, kódolás, többváltozós kapcsolatok

Kutató történészként gyakran kerülünk szembe azzal a problémával, hogy a szöveges adatok feldolgozása során a későbbi elemzés megkönnyítése végett az egyedi információinkat csoportosítani kell. Az így kapott csoportokat pedig kódolni szükséges. Ez gyakorlatilag a szótárak létrehozását jelenti. Az adatfeldolgozást előkészítő kutatócsoporti megbeszélések egyik fajsúlyos kérdése szokott lenni az adatkódolás elveinek előzetes kidolgozása. Ha a relációs adatbázisok lehetőségei felől tekintünk a kérdésre, ez a munkafázis, legalábbis a kezdet kezdetén, teljesen feleslegesnek minősíthető, hiszen a tábláinkba bármikor szabadon vehetünk fel újabb oszlopokat, ellátva ezáltal a rekordokat újabb információkkal. Nézzük meg ezt például a foglalkozások esetében.



11. ábra A foglalkozások rögzítésének lehetséges táblastruktúrája

Az ábrán láthatjuk, hogy a „Személy” tábla egy cellájában rögzítjük a foglalkozást. Ezt a forrás feldolgozásakor megtehetjük. Az egyedi foglalkozásokat külön azonosítóval látjuk el, és a foglalkozás táblában tároljuk. Minden egyes variációt külön-külön rögzíthetünk. Lehetnek ezek akár egymás szinonimái is, és ezáltal, ha fontos számunkra, akár a névhasználat történetisége is megjelenhet a rögzített adatainkban. Ebben az esetben tehát a „Foglalkozás” tábla szótártáblaként funkcionál. Egy másik munkafázisban magukról a foglalkozásokról kezdünk el részletesebb adatokat gyűjteni, fontos lehet számunkra, hogy ezek a konkrét foglalkozások milyen csoportokba rendeződnek. A 19. század végi adatgyűjtés esetén egymás után vesszük kezünkbe például a különböző népszámlálásokat, és rögzítjük az egyes foglalkozáshoz az adott évben használt statisztikai besorolást. Ezek a besorolások pedig a vármegye táblához hasonlatosan egy külön szótárban tárolódnak el. A foglalkozásra vonatkozó eredeti információnk, a pontos név tehát torzulás nélkül, a forrásban szereplő módon kerülhet be a rendszerbe. A csoportosítást, a kódolást pedig az adatfeltárás után, éppen a kutatás tapasztalatait is figyelembe véve végezhetjük el. A módszer nagy előnye, hogy nem jár információvesztéssel.

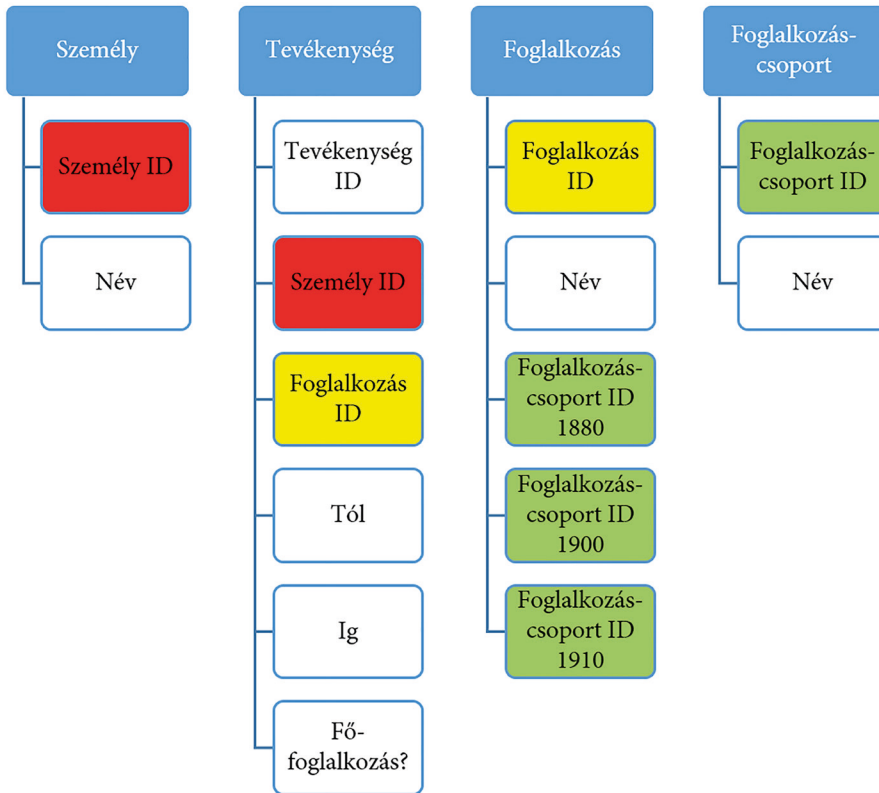
A fenti példát éppen amiatt használtam, mert köztudomású a társadalomtörténeti szakirodalomban, hogy a 19. századi magyar népszámlálások egymástól eltérő módon csoportosították a foglalkozásokat. Ez pedig megnehezíti a különböző időmetszetek összehasonlítását, a foglalkozásszerkezet átalakulási tendenciájának a kirajzolását.¹⁷ A fenti táblastruktúra azonban lehetőséget nyújt számunkra arra, hogy ugyanabból az adatsorból különböző elvek alapján készíthessünk lekérdezéseket, és ezeket az adatsorokat össze tudjuk hasonlítani egymással. Idővel pedig a finomodó elemzési szempontjainkat alkalmazva szükség esetén könnyen újra tudjuk számolni korábbi értékeinket.

A foglalkozás azonban egy olyan kérdéskör, melynek esetében egy másik problémával is szembetaláljuk magunkat: ugyanis ha adatainkat, hasonlóan fenti példánkhoz, egy adattáblában rögzítjük, ahol a „Személy” tábla szolgált erre a célra, akkor egy cellában csupán egy információt helyezhetünk el. A foglalkozás esetén dönteni kell, mit is rögzítünk. Mit teszünk azonban abban az esetben, ha egyértelműen nem dönthető el a kérdés, vagy egy személynek valóban több foglalkozása volt? Vizsgált személyünk például orvosként praktizált, és közben tanított egy egyetemen. Természetesen nyithatunk egy újabb oszlopot, ahol a másodlagos, harmadlagos értékeket rögzítjük, de ez csak a későbbi elemzéseket fogja nehezíteni. De hasonló probléma adódhat például az iskoláztatás helyének a rögzítése során vizsgált személyeink többsége egy helyen végzi a középiskoláit, néhányan két városban is megfordulnak, de egy szereplő életrajzában három különböző iskolahely fordul elő. A házastársak, a gyerekek rögzítése is ezt a problémát veti fel. A következőkben erre kívánunk egy lehetséges megoldást kínálni a foglalkozások példájánál maradva.

A foglalkozás tehát egy olyan információ, mely nem írható le egyértelműen egy adattal, egyszerűen több is jellemezheti az adott embert. Arra is van azonban bőven pél-

17 A problematikáról részletesen: Gyáni–Kövér 2003: 72–76.

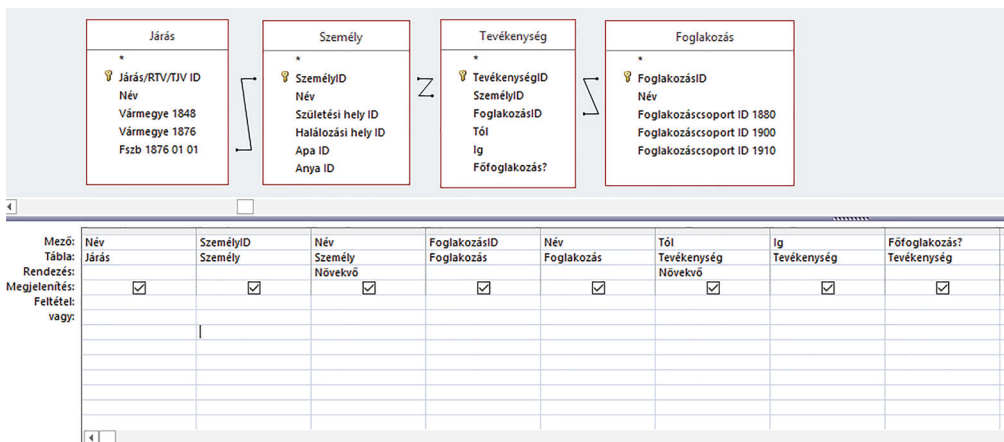
da, hogy idővel változik. Emiatt nem célszerű tehát a „Személy” táblában elhelyeznünk azt. Az előzőekben alkalmazott rögzítési rendszert tehát át kell alakítanunk. Az adat-rögzítés megoldását a következő ábrával szemléltethetjük.



12. ábra A tevékenységtípusok rögzítésének lehetséges táblastruktúrája

Az ábrán jól látható, hogy az átalakítást követően maguknak a foglalkozásoknak az adatrögzítése változatlan maradt, a „Személy” és a „Foglalkozás” tábla közé azonban bekerül egy olyan tábla, mely alkalmas a többszörös és időben változó kapcsolatok rögzítésére. Ez utóbbi táblát neveztem el „Tevékenység” táblának. A „Tevékenység” táblában egy emberhez több sort rendelhetünk, ezeket a sorokat látjuk el egyedi azonosítóval (TevékenységID), az időhatárok megjelölésével pedig biztosíthatjuk azt, hogy az egyes információk időbelisége is pontosan rögzíthető és követhető legyen.

Ha a fenti struktúrát alkalmazzuk az adatbázisban, akkor összetettebb kérdéseket is összeállíthatunk. A következő ábra egy olyan lekérdezést mutat be, melyben az összes járásunk, 1876. január 1-én szolgálatba lévő főszolgabíróinak az összes foglalkozását szeretnénk időrendben megkapni.



13. ábra Lekérdezés a mintaadatbázisban

A fenti ábrán már egy kicsivel több információt helyeztem el. Látható, hogy csak a szükséges táblákat vontam be a lekérdezésbe. A „Járás” tábla Fszb 1876 01 01 oszlopában található meg a főszolgabíró személy ID-je, ezt összekapcsoltam a „Személy” táblával. Ezt követően a „Személy” tábla személy ID-je biztosít kapcsolatot a „Tevékenység” tábla személy ID-jével. Innen pedig a foglalkozás ID jelent hidat a „Foglalkozás” tábla irányába. A feltett kérdés megválaszolásához szükségtelen táblákat nem használjuk fel, nem kérdeztünk ugyanis rá például a vármegyére, a személyes adatokra, vagy a foglalkozáscsoportokra.

Miután elkészítettük a lekérdezést, kiválaszthatjuk azokat az oszlopokat, melyeket meg szeretnénk jeleníteni. Ezeket szabadon variálhatjuk, sorrendjükről egyedileg dönthetünk. Jelen esetben a járás neve szerepel az első oszlopban, ezt követi a személy azonosítója és neve. Mivel kapcsolat van a „Járas” és a „Személy” tábla között, csak azokról kapunk információt, akik főszolgabíróként szerepelnek a „Járas” táblában. Ezen személyek foglalkozásait, azok időhatárait, és a főfoglalkozásra vonatkozó kérdésre adódó választ jelenítettem meg az eredménytáblában. Azt is meg lehet figyelni, hogy az adatokat többszörösen is rendezhetjük. A személyek alfabetikus sorrendben követik majd egymást, és egy személy foglalkozási adatai az időrendnek megfelelően jelennek meg, erre utal a két oszlopban beállított „növekvő” kifejezés. Ha a lekérdezést lefuttatjuk, akkor a következő adatsort kapjuk (a könnyebb értelmezhetőség miatt adatokkal töltöttem fel az eredménytáblát).

Név Járás	SzemélyID Személy	Név Személy	FoglalkozásID Foglalkozás	Név Foglalkozás	Tól Tevékenység	Ig Tevékenység	Főfoglalkozás? Tevékenység
Egri	2	Albert Albert	1	ügyvéd	1871.01.01	1895.12.31.	Igen
Egri	2	Albert Albert	2	segédtanár	1873.01.01.	1875.01.01.	Nem
Egri	2	Albert Albert	3	adjunktus	1875.01.02.	1880.01.01.	Nem
Egri	2	Albert Albert	4	docens	1880.01.02.	1890.01.01.	Nem
Egri	2	Albert Albert	5	egyetemi tanár	1890.01.02.	1895.12.31.	Nem
Szolnoki	1	Zoltán Zoltán	1	ügyvéd	1872.01.01	1886.12.31.	Igen

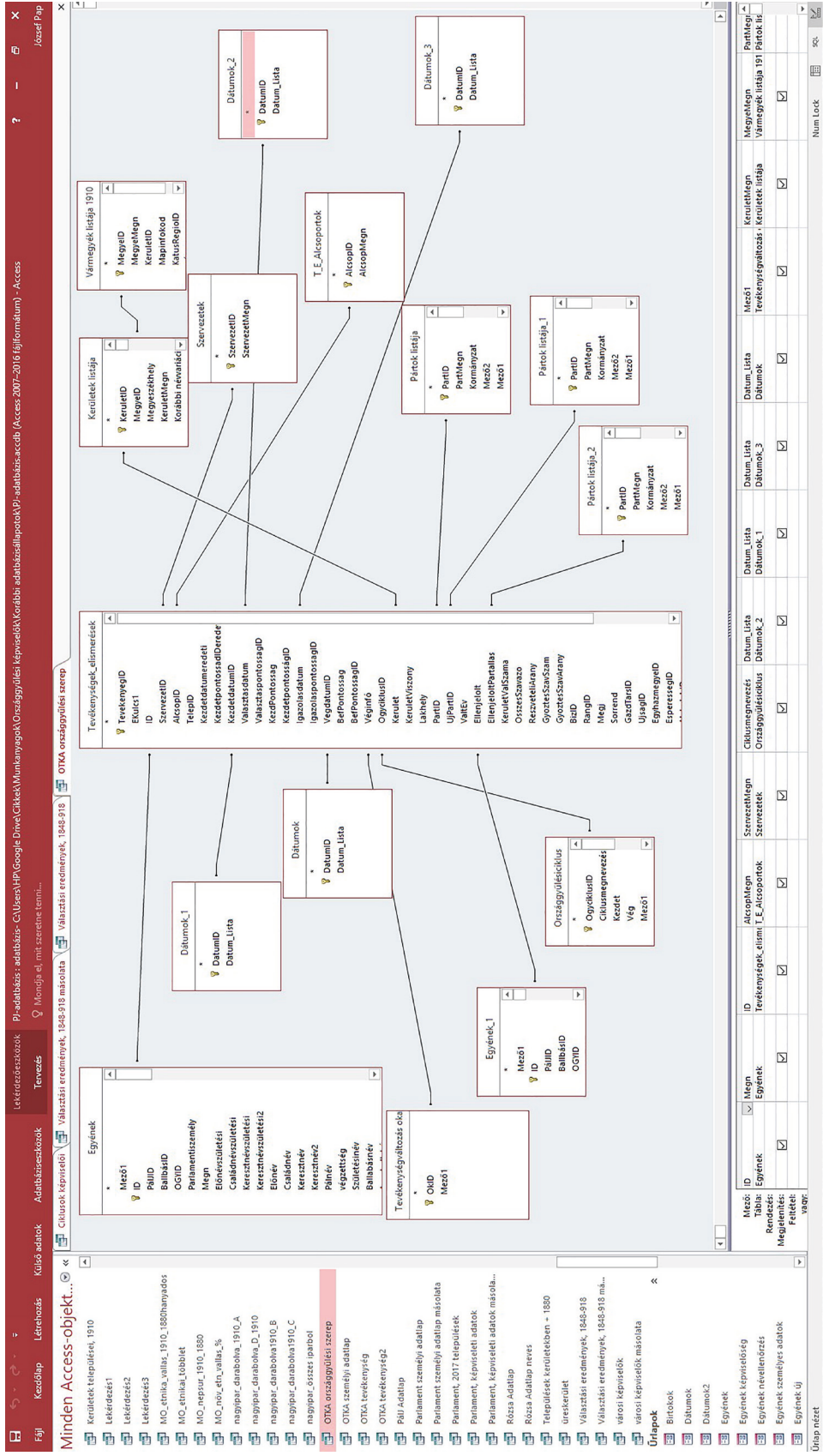
14. ábra A főszolgabírók foglalkozására vonatkozó lekérdezés eredménytáblája

A lekérdezés eredményeképpen tehát két különálló járás főszolgabírójának fiktív foglalkozási életpályáját kaptuk meg.

Ilyen, a fenti példának megfelelő táblastruktúrát bármire kidolgozhatunk, legyen az az oktatás vagy a társadalmi szerepvállalás, esetleg a családi kapcsolat. Bonyolultabb eljárással pedig minden, az egyénre jellemző információt egy táblában tárolhatunk. Ehhez nem kell más, mint hogy a tevékenységeket típusokba vonjuk össze, és olyan kategóriákat alakítunk ki, melyeket megfelelnek az egyéni életpályák egyes elemeinek. Ebben az esetben önálló tevékenységtípusok lehetnek például tanulmányok, a foglalkozások, a társadalmi szerepvállalás, a kitüntetés és elismerés stb. Az eljárás lényege, hogy egy táblában tárolva a típusazonosító segítségével elválasztjuk egymástól a különböző életpályaelemeket. Szükség esetén ezeket külön-külön lekérdezhetjük, de lehetőség van arra is, hogy egymást követően jelenítsük meg őket, megkönnyítve ezáltal az egész életpálya áttekintését.

A fenti példák jól illusztrálták, hogy adatbázisunk szabadon variálható táblázatok rendszeréből áll össze. Akár arra is van azonban lehetőségünk, hogy a szöveges dokumentumot helyezzük el rendszerünkben. Ehhez csak egy nagy mennyiségű szöveg tárolására alkalmas mezőt kell felvinni az egyik táblába. A Személyek táblán például életrajz elnevezéssel létrehozhatunk egy ilyet, ahol magát az adatrögzítés során feldolgozott életrajzot (vagy esetleg a kapcsolódó PDF, repozitóriumban elhelyezett állomány linkjét) is megőrizhetjük. Ha így teszünk, akkor az adatrögzítés igazából hasonlatossá válik a könyvtári katalógusrendszerekben használatos metaadatoláshoz.

Ha már évek óta fejlesztünk egy adatbázist, akkor igen bonyolult rendszer jön létre, melyből már összetett lekérdezések segítségével nyerhetünk adatokat. Erre vonatkozó példa látható a következő ábrán.



15. ábra Lekérdezés az általam használt adatbázisban

Az ábrán a kutatásom központi elemét jelentő kérdésnek, az országgyűlési szerepvállalásnak, az adatbázisomból való megjelenítését ábrázoló lekérdezés látható. Itt is megfigyelhetjük, hogy több táblát többször is összekötöttem a lekérdezés központi táblájával, a „Tevékenységek elismerések”-kel. Ez utóbbi rögzíti ugyanis a kategorizált tevékenységeket, melyek közül az egyik az országgyűlési szerepvállalás. A lekérdezésben többször szerepet kap a dátumokat, a településeket, a pártokat és a vizsgált személyeket tartalmazó tábla. A lekérdezés eredményeképpen megkapjuk az országgyűlési szerepvállalás adatait, melynek formája hasonlatos a főszolgabírók esetében bemutatott eredménytáblához, csupán jóval több adatot tartalmaz. Természetesen ebben az esetben is igaz, hogy ha a paramétereket másként állítom be, akkor a lekérdezés végeredménye más lesz, eredménye lehet akár az országgyűlési képviselők képzettségi adatait tartalmazó tábla is. Magukat a lekérdezéseket egyébként egymással is szabadon kombinálhatom, külön-külön listázhatom például az egyes pártok különböző időszakokban parlamentben jelen lévő képviselőinek képzettségi adatait. Ezek az önálló lekérdezések elmenthetők, ami nagyon előnyös, hiszen ha folyamatosan bővül a feldolgozott adattartalom, az új elemek egy újabb futtatás esetén beépülnek a korábbi eredményekbe.

Egy adatbázisrendszer az adatrögzítést is megkönnyíti a lekérdezések tervezéséhez használt grafikus felülethez hasonló eszközök segítségével. Ezek használatával a tábláink elé adatbeviteli űrlapokat szerkeszthetünk, így nem kell a kódok keresgetésével bajlódni, megteszi ezt helyettünk az űrlap automatizált adatbeviteli rendszere.

B. Leküzdendő akadályok

Adattábláinkat tehát a köztük keletkező kapcsolatok teszik relációs adatbázissá. Az ilyen adatbázis előállítása természetesen megkívánja a történésztől az informatikai eszköz ismeretét, azt a tudást, mely az adatbázis létrehozásához szükséges. De az adatbázis-építés „lelke” nem az informatikai megvalósításban rejlik, ennél sokkal fontosabb az a konkrét adatmodell, mely adatbázisunk mögött áll, melyet a források ismerete alapján tudunk felállítani.¹⁸ Ez pedig nem informatikai feladat, ehhez esetünkben egy történészre van szükség. Meg kell határoznunk ugyanis, hogy az elemzett kérdésről milyen információkkal rendelkezünk, és ezeket milyen struktúrával bíró táblázatokban tudjuk rögzíteni. Az így megalkotott rendszert lehet adatbázisba foglalni, és ezen utóbbi lépés az elsőhöz képest kevésbé lényeges. A történész számára azonban az informatikai létrehozás munkafázisa egy misztikus területnek tűnik. Az, aki nem tudja ilyen szinten használni a programot, az adatmodell megszerkesztése során megkereshet egy informatikában járatosabb kollégát, aki fizikai valóságában létrehozza neki a táblastruktúrát. Ehhez egyébként nem programozásra, hanem programhasználatra van szükség, melyhez jól hasznosítható segédkönyvek állnak rendelkezésre.

¹⁸ Az adatbázis és az adatmodell kapcsolatáról lásd részletesen Halassy 2000: 23–46.

kezésünkre.¹⁹ Azonban célszerű megbarátkoznia minden adatbázist használni kezdő kollégának a gondolattal, hogy előbb-utóbb el kell jutni az önálló programhasználat fokára. Ugyanis adatbázisunk dinamikusan fog fejlődni, és nagyban nehezíti munkánkat, ha minden egyes részproblémával az informatikusunkat kell megkeresnünk.

Az adatbázis létrehozása tehát nem az informatikus feladata, hanem egy informatikai ismeretekkel bíró, vagy informatikai tudással rendelkező segítővel együtt dolgozó történészé. Az adatbázis-építés ugyanis nem nélkülözi a klasszikus történetírói módszereket, a forrásfeltárást, a forráskritikát. Egy olyan eszközként is tekinthetünk a relációs adatbázisra, mint amely modern formában jeleníti meg a klasszikus cédulázást, ahol az egyes kártyák, a rajtuk elhelyezett információkkal, gyakorlatilag egy manuális adatbázist jelentettek. Míg azonban korábban a történész manuálisan keresgette és rendezgette céduláit, most az adatbázisának tartalmát rendezi össze lekérdezések segítségével. Ezt azonban sokkal gyorsabban és hatékonyabban tudja ma már megtenni.

Ezt követően lehet feltenni azt a kérdést, hogy vajon mivel adott volna többet 1997 és 2003 között egy adatbázis használata számomra? Úgy látom, hogy az adatbázis alkalmazásának egyik legfontosabb előnye az lett volna, hogy minden egyes hivatalviselő csak egy alkalommal került volna felvételre. Nagy problémát jelentett ugyanis számomra utólag az, hogy a tisztviselőket, mivel a forrásadottságoknak megfelelően önálló megyei listákból dolgoztam, megyénként gyűjtöttem össze. Az egyező személyek kiszűrése jelentős nehézséget okozott, és ezt követően több táblázatban kellett egyenként kijavítani az adatokat. Ez elkerülhető, ha a „Személy” táblában található meg minden szereplő, szükség esetén utólag könnyebben elvégezhető a duplázódó egyéneknek és tevékenységüknek az összevonása. Egyébként a dupla rögzítés is könnyebben kiküszöbölhető, hiszen van arra mód, hogy olyan űrlapot készítsünk, ahol a rögzítés előtt ellenőrizhetjük, hogy az adott név szerepel-e már a listánkban.²⁰

Problémát jelentett korábban továbbá az is, hogy az egyes politikai időszakok esetében megyénként kellett összeszámolgatni a különböző adatokat, áttekinteni az életpályamodelleket. Ez az adatbázisban a tevékenységek tábla egyedi adatsorainak utólagos kódolásával oldható meg, amit minden személy esetében csupán egyszer kell elvégezni. Ezt követően pedig egy olyan lekérdezést kell összeállítani, mely azon túl, hogy megyénként összesíti az adatokat, tekintettel van a különböző politikai időszakokra is. Mivel a lekérdezések elmenthetők, ha valamely adatról kiderül, hogy téves, vagy újabb, korábban nem ismert információra bukkanunk, nem kell minden esetben újraszámolni értékeinket, elég csupán újra lefuttatni a lekérdezést. Másik fontos előny tehát a nagyobb

19 Magam a kezdeti lépéseket a Dummies sorozat Access 2003-as kötetének segítségével sajátítottam el, mely, mint ahogy a címe is mutatja, nem éppen az informatikában járatos személyek számára íródott. Időközben azonban különböző programverziókhoz újabb kötetek jelentek meg, jelenleg a 2019-es verzió érhető el (Ulrich, Laurie A. – Cook, Ken 2019: Access 2019 For Dummies. Hoboken, New Jersey.). A kézikönyvek mellett azonban nagy segítségünkre lehet a Youtube is, ahol könnyen találhatunk olyan videókat, melyek az adott program használatába vezetik be a különböző szintű tudással rendelkező felhasználókat (pl.: <https://www.youtube.com/playlist?list=PLXP4h6BgzLN1xdJEAhsj8kDDCL-if1pGj>; <http://oktatas.tanarurkerem.com/>).

20 A beviteli űrlappal kapcsolatban lásd: Nagy 2019: 81–86.

fokú pontosság. Mivel a számításokat nem manuálisan végezzük, kisebb a tévedés lehetősége, az ellenőrzés miatt pedig nem kell minden matematikai műveletet többször elvégezni. Ebből következik, hogy különösen ki kell emelni az időtényezőt. Az adatbázis felépítése természetesen jelentős időt vesz igénybe, de az adatrögzítés már sokkal egyszerűbben zajlik. A lekérdezések összeállítása néhány perc csupán, ezt követően pedig szinte azonnal eredményeket kapunk. A disszertációm készítése során csak ez utóbbi fázis heteket vett igénybe, hiszen teljesen manuálisan kellett végrehajtanom a számításokat.

Van azonban még egy hatalmas előny, ez pedig a szerves építkezés lehetősége, hiszen alapvetően ekkor is emberekről gyűjtöttem adatokat, akik ebben a kutatásban tisztviselőként kerültek a vizsgálat középpontjába. Ugyanebben az adatbázisban azonban el lehet helyezni olyan személyeket, akik például országgyűlési képviselők voltak. Mindkét kérdéskör – vármegyei hivatal, országgyűlési képviselő – egy-egy tevékenység típusnak, életpályaelemnek tekinthető. A különböző kutatási programok során feltárt személyek között egyrészt előbb-utóbb személyazonosság is fel fog lépni, másrészt egy részük rokoni kapcsolataik révén is össze fog kapcsolódni egymással. A korszak szereplőire vonatkozó információink tehát, a különböző megközelítésmódok ellenére, egymásra épülnek, egymással relációs kapcsolatba kerülnek. Adatbázis nélkül egyik kutatást követi a másik, és csak korlátozott memóriánk teremt kapcsolatot a feldolgozott emberek tömegei között.

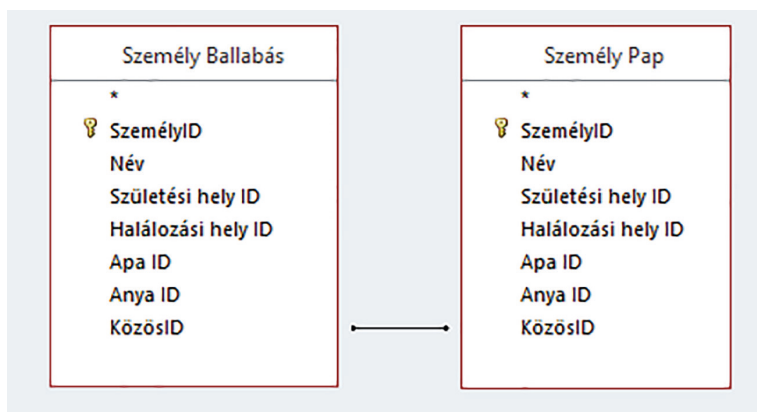
Fontos kérdés azonban az is, hogy mi akadályozza az adatbázis utólagos alkalmazását, ha korábbi adataim táblázatokban tárolva természetesen a mai napig megvannak. Ennek alapvetően két oka van. Az egyik az adatkonvertálás problémája, amit automatizálni nem nagyon lehet. Gyakorlatilag egyszerűbb újrakezdeni az adatok bevitelét, mint azzal kísérletezni, hogy a különböző táblatípusokban Wordben, Excelben tárolt adatokat valahogy automatikusan kódoljuk. Természetesen erre is van mód, de ez már valóban nagy informatikai háttértudást igényel. A másik, talán fontosabb ok, maga az adatok kódolása, melyet annak idején én is az adatrögzítés során végeztem el. Az információ rögzítésekor nem minden esetben őriztem meg az eredeti alakot (pl. a foglalkozás pontos megnevezését), hanem valamilyen kategóriába soroltam be azt, gyakorlatilag ez azt jelenti, hogy újra kezembe kell venni az eredeti forrást. Ha korábbi adatainkból akarunk adatbázist építeni, a legritkább kivételtől eltekintve, vissza kell térni a már feldolgozott anyagokhoz, jól felépített adatbázisok egymástól szeparált adattáblákból nem keletkeznek. A befektetett energia azonban megtérül.

Összefoglalásképpen nem lehet azonban eléggé hangsúlyozni, hogy milyen nagy figyelmet kell szentelni az adatbázis megtervezésére. Ugyanis egy rosszul kialakított táblastruktúra esetén, az adatrögzítés során olyan problémákkal kerülhetünk szembe, melyek csak az adatbázis áttervezésével oldhatók meg. Egy rosszul megválasztott adattípus is komoly adatvesztéssel járhat. El kell kerülnünk az adatredundanciát, a felesleges adatrögzítést, ügyelni kell arra, hogy hasonló adattípusokat ne különböző táblákban rögzítsünk. Lényeges továbbá az is, hogy a csoportképzést minden esetben az egyedi adatok rögzítése után végezzük el. Hiszen az aggregált értékekből az egyedi jellemzők már nem állíthatók helyre.

C. Adatbázisok relációs kapcsolatban

Lehetőségeink azonban itt nem fogynak el, hiszen adatbázisunk más adatbázissal is összekapcsolható. Ez gyakran még abban az esetben is megtehető, ha nem azonos témában végzik a kutatást az adatbázisok készítői. Ha találunk érintkező pontokat, akkor ez a kapcsolat létrehozható. Most röviden két olyan kutatási eredményről szólok, mely a relációs adatbázisok összekapcsolásából származik, anélkül nem is jött volna létre.

Az első példa személyekhez kapcsolódik, a törvényhozás két házának, a képviselőháznak és a főrendi háznak a tagjait köti össze. Ebben az esetben Ballabás Dániel és jómagam adatgyűjtése szerepel. A Ballabás Dániellel folytatott együttműködés 2007-es kezdetét követően két külön adatbázist építettünk. A két különálló rendszer azonban könnyedén összekapcsolható, ugyanis bizonyos alaptáblák – idő, település, nemesi famíliák stb. – közös tartalmúak, közösen fejlesztettek. Emellett a legfontosabb információk tárolásának módja – kapcsolatok, tevékenységcsoportok, forrásmegjelölés – is azonos. A kapcsolatot jelen esetben a személyek adják. A saját kutatásunk során a külön-külön felvett egyéneket, a vizsgálat alanyait ugyanis közösen beazonosítottuk, beazonosítjuk, így a személyi adatokat tartalmazó adattábláink egyénszinten összekapcsolódnak. Ezáltal a két adatbázisban található, különböző tevékenységtípusra vonatkozó információk összeadódnak. Emellett lehetőség van arra is, hogy a két adatgyűjtésben szereplő, azonos tevékenységhez kapcsolódó, de egymásnak ellentmondó adatokat kiszűrjük, ellenőrizzük.



16. ábra: Két adatbázis kapcsolata közös azonosítók segítségével

Saját kutatásom során a dualizmus korának országgyűlési képviselőit gyűjtöttem össze. Az első népképviseleti választástól 1918. november 16-ig 3281 fő szerepelt a magyar parlamentben, valamint 29 esetben ismeretlen a képviselő személye. Ballabás Dániel adatai alapján ugyanezen intervallumban 1998 főrendről tudunk. Ez a két csoport alkotja az országgyűlési résztvevők körét, mely azonban természetesen nem azonos

a két halmaz összegével, uniójával. Valójában csupán 4860 főről beszélhetünk, ugyanis 421 személy, tehát a teljes sokaság 8,6%-a főrend és képviselő is volt egyben.

A két egyesített lista nagy hozadéka ez a pontos, személyszinten azonosított politikusi kör, mely kutatásaink metszeteiként jelenik meg. A tudományos kutatások általában egy-egy csoportra koncentrálnak, de az azonos alapon és rendszerben gyűjtött adatok lehetővé teszik számunkra, hogy a különböző vizsgálatok eredményeként keletkező halmazok átfedéseit, metszeteit is vizsgálni tudjuk, vagy másként fogalmazva, a részek helyett az egészre tekinthessünk. A politikai elit életpályaelemzése tehát csak e két munka összekapcsolásával végezhető el, hiszen az átfedések figyelembevétele nélküli vizsgálat fals eredményre vezetne. A politika társadalomtörténete szempontjából különösen érdekes ez a csoport, hiszen vizsgálatuk lehetőséget nyújt például olyan kérdések megválaszolására, hogy milyen hatással volt az 1885-ös főrendi házi reform az egyén politikai karrierjére? Mélyebben elemezhetővé teszi továbbá a főnemeselek képviselőházi jelenlétének problematikáját.

Hasonló alapon épült fel parlamentarizmustörténeti kutatócsoportunk adatgyűjtése is, hiszen ugyanarra a sokaságra végezhetünk el különböző adatgyűjtéseket. Ballabás Dániel, Pál Judit és magam a politikai tevékenységet dolgoztuk fel, Gerhart Péter az oktatási viszonyokat, Sidó Zsuzsa az emlékezetet, Szendrei Ákos a foglalkozást, Tóth-Barbalics Veronika pedig a tudományos tevékenységet és társadalmi szerepet. Ezek a különálló adattáblák a vonatkozó személyek alapján kapcsolódnak össze, és így a részekből összeáll a politikus pályaképe.

Kevésbé egyértelmű azonban olyan adatgyűjtések összekapcsolása, mint amelyre a következő példa vonatkozik. Demeter Gábor vezette kutatócsoport „Térinformatikai rendszer kiépítése Magyarország és az Osztrák–Magyar Monarchia történetének tanulmányozásához” című projektjében Magyarország területére vonatkozó sokrétű adatgyűjtést hajtott végre, rögzítette többek között az 1910-es népszámlálás statisztikai adatai is. Demeter Gábor adatait a települések segítségével lehetett összekapcsolni saját adatbázissal. Jómagam a településekkel csupán áttételesen foglalkoztam, természetesen fontosak voltak számomra is, hiszen a képviselők életrajzához kapcsolódó helyként rögzítettem őket. A politika színtere azonban a választókerület volt. Azonban mivel a történeti Magyarország településeit a választókerületekhez rendeltem, így a két adatbázis között kialakult a szakmailag is releváns kapcsolat. A települések statisztikai adatai ugyanis választókerületi szinten összegeezhetők, így tudom megvizsgálni például a politikai jelenségek és a népszámlálási adatok közötti kapcsolatot. Olyan kérdéseket tehetek fel, hogy volt-e kapcsolat a kerület felekezeti adottságai és a képviselő felekezeti hovatartozása között, vagy adódott-e valamilyen politikai magatartásforma a lakosság felekezeti sajátosságaiból.²¹ Lehetőség volt arra is, hogy a kerületi etnikai

21 Pap 2016: 337–359. Kifejezetten érdekes a protestáns magyar kerületek ellenzéki karakterének és a katolikus magyar kerületek vegyes jellegének az összevetése ugyanezen területek foglalkozásszerkezeti adataival. Ebből az a következtetés vonható le, hogy míg a protestáns, döntően magyar kerületek dominánsan ellenzékiek voltak (Erdély kivételével), addig a katolikusok abban az esetben kerültek ebbe a kategóriába, ha ott a foglalkozásszerkezeten belül a gazdák aránya az átlagnál magasabb volt. A po-

viszonyok és a választói aktivitás kapcsolatát is vizsgáljam.²² Demeter Gábor kutatócsoportja pedig a választókerületek esetében a tudott olyan kérdéseket feltenni, mely a lakosság gazdasági szerkezete, etnikai adatai és a census mértéke között keresett kapcsolatot. Így tudta Demeter Gábor cáfolni azt az általánosan elfogadott tételt, hogy a nemzetiségi területeken az arányaiban alacsonyabb földjövedelem és census segítségével juttatták kedvezőbb helyzetbe a kormányhoz hűségesebb választókat.²³ Valamint felhasználhatták térképeim adatait a területi egyenlőtlenségek politikai következményeinek vizsgálatára is.²⁴

A relációs adatbázisok tehát sok esetben maguk is összekapcsolhatók, mintegy hálózatba rendezhetők, és ezen kapcsolatok segítségével olyan kutatási eredmények érhetők el, melyekre önállóan egyik adatgyűjtés sem képes. Ebben az esetben nemcsak arról van szó, hogy két adatsort összeadunk, hanem olyan horizontot nyitunk, ami korábban nem létezett. Választókerületenként összesített községoros adatokat például hiába keresünk a statisztikai irodalomban, ezt a 2010-es években hoztuk létre.

Fontos azonban újra kihangsúlyozni azt is, amire az első sorokban már utaltam. Az adatbázisépítés megtanulásával az ember olyan kompetenciához jut, mely nagymértékben hasznosítható az élet egyéb területein is. Az Eszterházy Károly Egyetem Történelemtudományi Intézetnek jelenlegi vezetőjeként igen jelentős adminisztratív munkát kell végezni, azt hiszem, ezzel nem vagyok egyedül, minden olyan ember, aki vezető feladatot vállal, szembesül ezzel. Az adminisztráció különböző típusaira azonban szintén kidolgoztam egy-egy adatbázist, melynek segítségével viszonylag rövid időn belül tudom produkálni a megfelelő típusú táblázatot (a táblázatoknak ugyanis az a sajátja, hogy évről évre változnak). Az itt felszabaduló munka, pedig a kutatásra fordított időben hasznosulhat. Egy adatbázist használó történész tehát a munkaerőpiacon is jól hasznosítható tudással, gondolkodásmóddal bír. Az adatbázissal való munka azonban erre vonatkozó affinitást is igényel, melynek megléte nélkül szinte felesleges a módszerrel kísérletezni. Úgy gondolom azonban, hogy egy kis nyitottsággal, egy új kutatás kezdetekor célszerű belevágni a dolog megismerésébe. Különösen igaz ez a pálya kezdetén álló fiatalok esetében.

litikát tehát nagymértékben befolyásolhatta 1848, a földhöz való hozzájutás élménye, melyet Kossuth Lajoshoz és a rá hivatkozó pártokhoz kapcsolhattak a dualizmus korának paraszti választói.

22 Pap 2019: 163–189. Különösen érdekes a szlovákság választási szereplése. A nyugati területeken, ahol a Szlovák Nemzeti Párt jelöltek tudott állítani, jóval magasabb választói aktivitás tapasztalható. A keleti, szegényebb kerületek esetében sokkal gyakoribb a választások elmaradása, vagy az érdemi eséllyel bíró kihívó nélküli választás. A két terület különbsége nem csupán felekezeti jellegű volt, hanem a jövedelmi viszonyok is eltértek egymástól. A szegényebb keleti végeken jóval alacsonyabb a választói aktivitás, mint a jobbmódú nyugati kerületekben.

23 Demeter 2018a: 42–43.

24 Demeter 2018b: 181., 203.

Hivatkozott irodalom

- Ballabás Dániel 2019: Családfákon innen és túl. Genealógiai kapcsolatok detektálása a hálózatok segítségével. In: Ballabás Dániel (szerk.): Hagyományos források, új megközelítések. A digitalizáció kínálta lehetőségek a történeti kutatásokban. Eger.
- Demeter Gábor 2018a: Kiegyensúlyozott fejlődés? Az agrárjövödelmek és adóterhek területi mintázata és változása a dualizmus kori Magyarország területén. In: Demeter Gábor – Szulovszky János (szerk.): Területi egyenlőtlenségek nyomában a történeti Magyarországon. Módszerek és megközelítések. Budapest – Debrecen.
- Demeter Gábor 2018b: Perifériák és fejlett régiók a történeti Magyarországon (1890–1910). In: Demeter Gábor – Szulovszky János (szerk.): Területi egyenlőtlenségek nyomában a történeti Magyarországon. Módszerek és megközelítések. Budapest – Debrecen.
- Gyáni Gábor – Kövér György 2003: Magyarország társadalomtörténete a reformkor végétől a második világháborúig. Budapest.
- Halassy Béla 2000: Adatmodellezés elmélet és gyakorlat. Budapest.
- Nagy Andor 2019: A digitalizált alkalmi írások kutatásának lehetőségei informatikai eszközökkel. In: Ballabás Dániel (szerk.): Hagyományos források, új megközelítések. A digitalizáció kínálta lehetőségek a történeti kutatásokban. Eger.
- Pap József 2003: Magyarország vármegyei tisztikara a reformkor végétől a kiegyezésig. Szeged
- Pap József 2007: „Két választás Magyarországon”: az országgyűlési képviselők társadalmi összetétele a 20. század első éveiben. *Aetas*, (22) 1.
- Pap József 2014: Tanulmányok a dualizmus kori magyar parlamentarizmus történetéből. Budapest
- Pap József 2016: Vallás és politikum a dualizmus korában: A felekezeti kérdés és a népképviselő az Osztrák–Magyar Monarchia Magyarországon. In: Peter, Kónya (szerk.): Rekatolizáció és a magyar társadalom a XVII–XVIII. században (Rekatolizáció a Uhorská spolocnost a 17. – 18. storoci). Eperjes.
- Pap József 2019: A dualizmus kori parlamentarizmus térbelisége. In: Gál Máté – Péterffy Gergely (szerk.): 1918: vég és kezdet (Régióink története I.). Eger.
- Rózsa Sándor 2019: Térinformatika és történettudomány. In: Ballabás Dániel (szerk.): Hagyományos források, új megközelítések. A digitalizáció kínálta lehetőségek a történeti kutatásokban. Eger.
- Tuza Csilla 2016: Állami szolgálat, közszolgálat a 19. század közepén. Egy készülő abszolutizmuskori személyügyi adatbázis. *Századok* (150) 5.